What is claimed is:

1. A computer-implemented method of identifying key images in a document, comprising:

extracting one or more document keywords from the document considered important in describing the document;

collecting one or more images associated with the document including information describing each image;

generating a proximity factor for each image collected from the document and each document keyword that reflects the degree of correlation between the image and the document keyword; and

determining the importance of each image according to an image metric that combines the proximity factors for each document keyword and image pair.

2. The method of claim 1 further comprising presenting the images within the document determined to be important on a display device.

3. The method of claim 1 further comprising:

ordering the document keywords according to an ordering criterion; and

weighting the proximity factor associated with each document keyword and image pair based on the order of the document keyword.

4. The method of claim 3 wherein the frequency that each document keyword appears in the document determines the ordering criterion used to order the document keywords.

5. The method of claim 3 wherein the ordering criterion orders the document keywords according to their relationship with the subject matter of the document.

6. The method of claim 1 wherein generating a proximity factor further comprises identifying image text used to describe each image.

7. The method of claim 6 wherein identifying the image text further comprises:

scanning a bit-mapped representation of the image for text information; and

converting the bit-mapped representation of the text information into image text.

8.   The method of claim 6 wherein the images include metadata having text information and identifying the image text further comprises:

searching the metadata information associated with the image for text describing the image.

9.   The method of claim 8 wherein the metadata information is compatible with hypertext markup language (HTML).

10.   The method of claim 6 wherein generating the proximity factor further comprises lexically analyzing the image text associated with each image and each document keyword to determine the degree of correlation between an image and a document keyword.

11.   The method of claim 6 wherein generating the proximity factor further comprises performing a phonetic comparison between the image text associated with each image and each document keyword to determine the degree of correlation between an image and a document keyword.

12.   The method of claim 1 wherein generating the proximity factor further comprises:

identifying the location of the image in the document; and measuring the distance between the image in the document and a document keyword.

13.   The method of claim 11 wherein generating the proximity factor further comprises:

determining the correlation between each document keyword and an image according to the distance between the document keyword and the image.

14.   A computer apparatus for identifying key images in a document, comprising:

a processor for executing instructions;

memory having instructions when executed on the processor that extract one or more document keywords from the document considered important in describing the document, collect one or more images associated with the document including information describing each image, generate a proximity factor for each image collected from the document and each document keyword that reflects the degree of correlation between the image and the document keyword, determine the importance of each image according to an image metric that combines the proximity factors for each document keyword and image pair.

15. The apparatus of claim 14 further comprising instructions when executed on the processor that present the images within the document determined to be important on a display device.

16. The apparatus of claim 14 further comprising instructions when executed on the processor that order the document keywords according to an ordering criterion and weight the proximity factor associated with each document keyword and image pair based on the order of the document keyword.

17. The apparatus of claim 16 further comprising instructions when executed on the processor that use the frequency that each document keyword appears in the document to determine the ordering criterion used to order the document keywords.

18. The apparatus of claim 16 further comprising instructions when executed on the processor that uses an ordering criterion that orders the document keywords according to their relationship with the subject matter of the document.

19. The apparatus of claim 14 wherein the proximity factor is generated by instructions executed on the processor that identify image text used to describe each image.

20. The apparatus of claim 19 further comprising instructions when executed

on the processor that scan a bit-mapped representation of the image for text information and convert the bit-mapped representation of the text information into image text.

21.   The apparatus of claim 19 wherein the images include metadata having text information and identifying the image text further comprises instructions when executed on the processor that search metadata information associated with the image for text describing the image.

22.   The apparatus of claim 21 wherein the metadata information is compatible with hypertext markup language (HTML).

23.   The apparatus of claim 19 wherein generating the proximity factor further comprises instructions when executed on the processor that lexically analyze the image text associated with each image and each document keyword to determine the degree of correlation between an image and a document keyword.

24.   The apparatus of claim 19 wherein generating the proximity factor further comprises instructions when executed on the processor that perform a phonetic comparison between the image text associated with each image and each document keyword to determine the degree of correlation between an image and a document keyword.

25.   The apparatus of claim 19 wherein generating the proximity factor further comprises instructions when executed on the processor that identify the location of the image in the document and measure the distance between the image in the document and a document keyword.

26.   The apparatus of claim 25 wherein generating the proximity factor further comprises instructions when executed on the processor that determine the relevance of each document keyword to an image according to the distance between a document keyword and the image.

27.   An apparatus for identifying key images in a document, comprising:

- 14 -

means for extracting one or more document keywords from the document considered important in describing the document;

means for collecting one or more images associated with the document including information describing each image;

means for generating a proximity factor for each image collected from the document and each document keyword that reflects the degree of correlation between the image and the document keyword; and

means for determining the importance of each image according to an image metric that combines the proximity factors for each document keyword and image pair.

28.   The apparatus of claim 27, further comprising:

means for ordering the document keywords according to an ordering criterion; and

means for weighting the proximity factor associated with each document keyword and image pair based on the order of the document keyword.